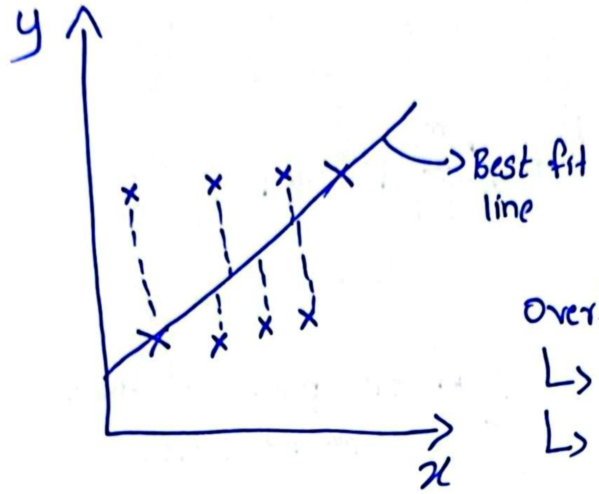
RIDGE REGRESSION (L2 Regularization)



Overfitting
↳ Traning Data → High Accuracy → Low Bias
↳ Test Data → Low Accuracy → High Variance

To reduce overfitting, we can use Ridge Regression

Cost Function

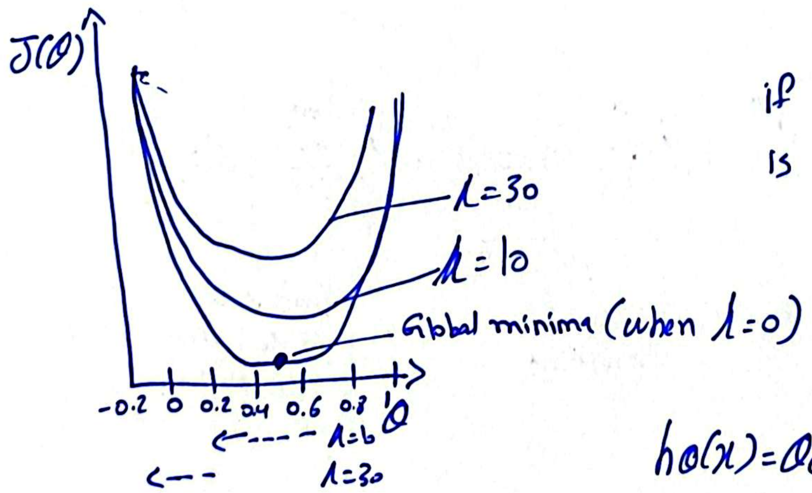
$$\frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^i - y^i\right)^2 + \underbrace{\lambda}_{\text{hyperparameter}}\sum_{i=1}^{m}(\text{slope})^2$$

if $\lambda = 1$

$$h_\theta(x) = \theta_0 + \theta_1 x_1$$

$$= 0 + 1[\theta_1]^2$$

$$> 0$$



if λ is increasing, slope is descreasing

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

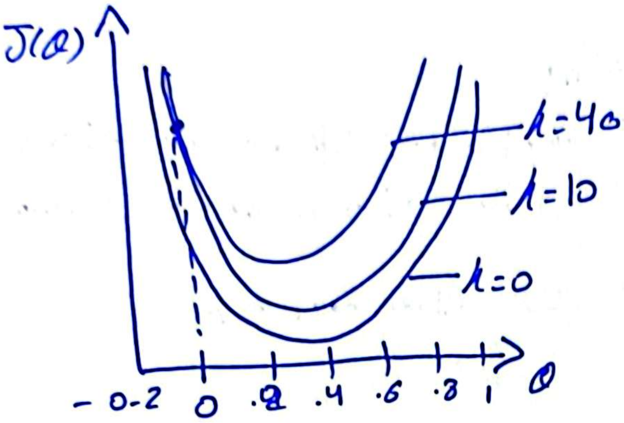
$$= \underset{0.34}{\cancel{0}} + 0.52 x_1 + 0.48 x_2 + 0.24 x_3$$

After apply ridge regression $= 0.34 + 0.40 x_1 + 0.38 x_2 + 0.14 x_3$

LASSO REGRESSION (L1 Regularization) → Feature Selection

Cost Function

$$\frac{1}{2m}\sum_{i=1}^{m}\left[h_\theta(x)^i - y^i\right]^2 + \lambda\sum_{i=1}^{m}|\text{slope}|$$



After one point of time, θ will become zero.

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

$$= 0.52 + 0.65x_1 + 0.72x_2 + 0.34x_3 + 0.12x_4$$

Lasso Regression: $0.52 + 0.65x_1 + 0.60x_2 + 0.14x_3$ $\quad \therefore 0x_4$ removed

Elastic Net (L1 and L2)

Cost Function

$$\frac{1}{2m}\sum_{i=1}^{m}\left[h_\theta(x)^i - y^i\right]^2 + \underbrace{\lambda_1\sum_{i=1}^{m}(\text{slope})^2}_{\text{reduce overfitting}} + \underbrace{\lambda_2\sum_{i=1}^{m}|\text{slope}|}_{\text{feature selection}}$$